Advances in hierarchical forecasting

Forecasting Hierarchies of Products and Market Segments

Nikolaos Kourentzes

Lancaster University Management School

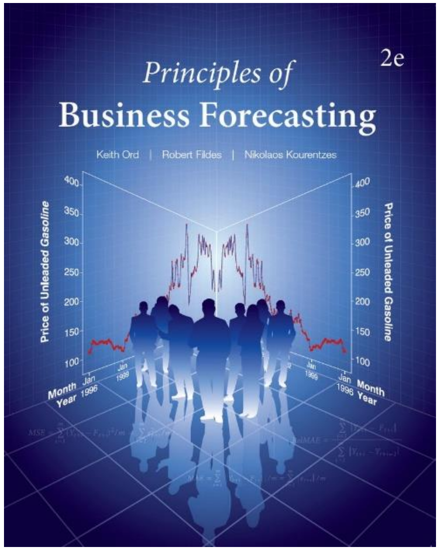
Centre for Marketing Analytics and Forecasting

13/04/2018

LUMS Marketing Analytics and Forecasting

Lancaster University Management School

Nikolaos Kourentzes



Associate Professor at Dept. Management Science
Lancaster University Management School, UK
Lancaster Centre for Marketing Analytics and Forecasting

Editorial Board of the International Journal of Forecasting, the leading academic journal in the field of forecasting

Research interests and consulting experience in various fields of forecasting, including:

- Business Forecasting and Demand Planning
- Promotional Modelling and Retailing and Marketing Analytics
- Artificial Intelligence
- Supply Chain Forecasting and Bullwip effect

Long experience in applied research projects with industry in various sectors, including: retailing, FMCG manufacturers, pharmaceuticals, media, energy, call centres among others.

Research blog: http://nikolaos.kourentzes.com
Book: Principles of Business Forecasting, 2017, 2nd edition, Wessex Press Publishing

Agenda

1. What is hierarchical forecasting?
2. Cross-sectional hierarchies
3. Temporal hierarchies

What is hierarchical forecasting?

Often forecasting problems exhibit a natural hierarchical structure. For example:

- Product with variants
- Products within product groups
- Market segments and geographical segments
- Different channels of distribution
- Services that share common resources (e.g. call centres)
- etc.

In such cases we can employ the so called “hierarchical forecasting” methods. The main objective of such approaches is to **ensure that forecasts are consistent across levels of the hierarchy.**

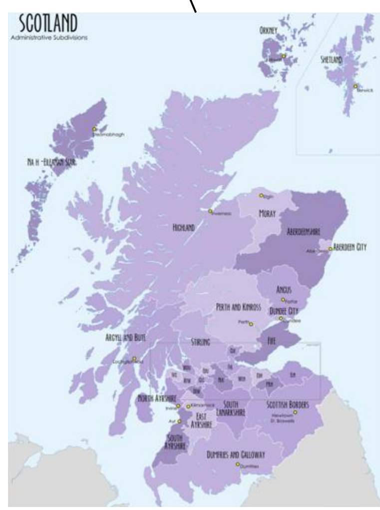
- Total country sales are consistent with sales in sub-regions, etc.

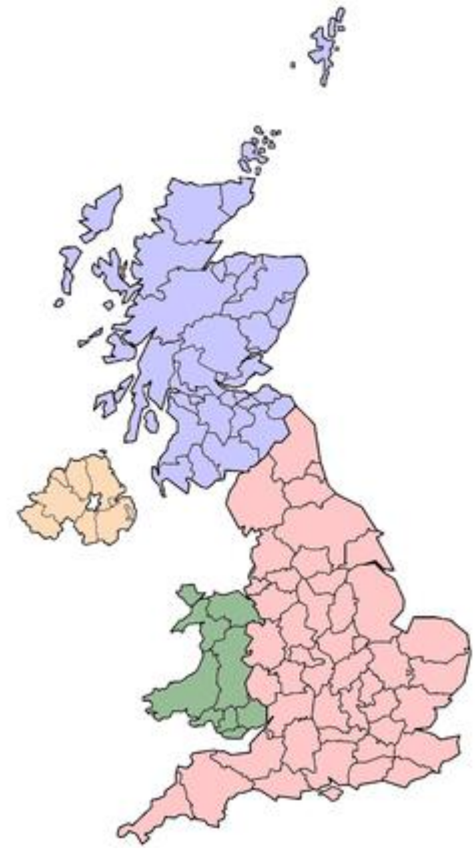
What is hierarchical forecasting?

As an example we can visualise the forecasting problem as follows:

UK

England | Northern Ireland | Scotland | Wales

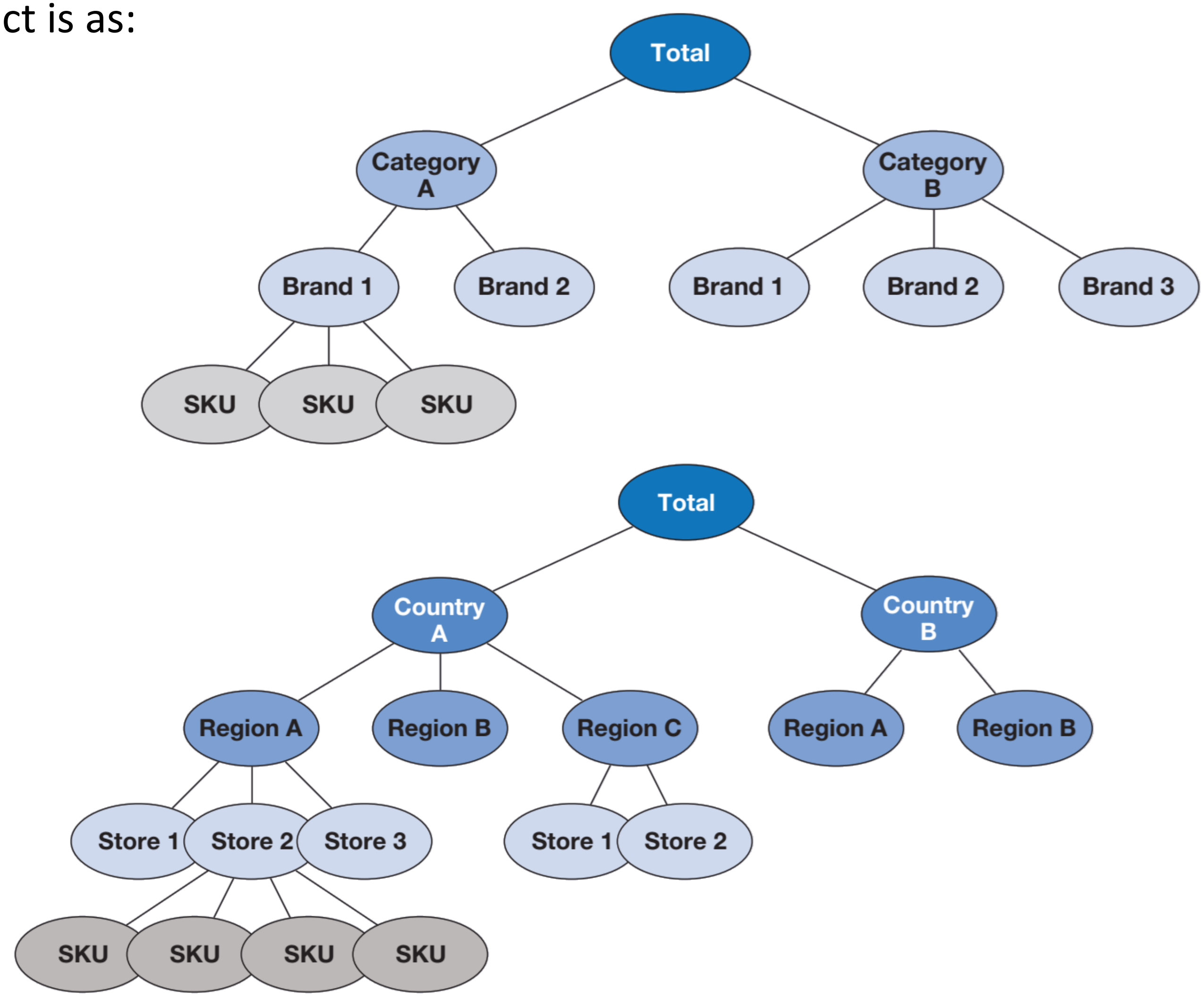




What is hierarchical forecasting?

Or more generally abstract is as:

Our hierarchies can have as many levels as we want, driven by the business.



Hierarchical and Grouped series

Note that in the previous examples we assumed that there was one way to get from the lowest level to the highest level, i.e. a single **hierarchy**.

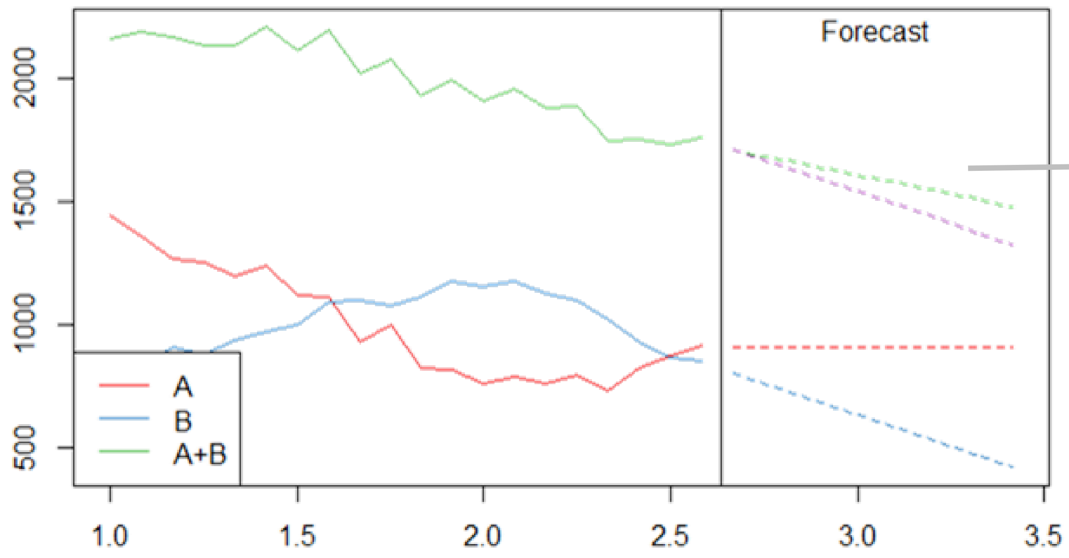
This is not generally true, as there may be many ways to construct the hierarchies, for example:

- SKU → Product group → Total
- SKU → Store → Total
- SKU → Country → Total
- etc.

We can represent all possible pathways from the disaggregate data to the top level aggregate data using the so called **grouped** time series.

The forecast consistency problem

Suppose we have to forecast two items A and B, which are variants of the same product.



Reconciling this difference imposes the aggregation constraint, and will force changes to the forecasts of A and B.

Hierarchical and Grouped series

An example from a policy problem, managing unemployment is as follows:

- **Sixteen** unemployment time series across the following dimensions:
  - Age {15-24; 25 and above}
  - Country {Denmark; Finland; Norway; Sweden}
  - Gender {Female; Male}

- From these we can construct multiple hierarchies, resulting in 29 unique aggregate series (16 + 29 = 45 series in total).

	Top Level	Level 1	Level 2	Level 3
Hierarchy 1	Total	Country	Gender	Age
Hierarchy 2	Total	Country	Age	Gender
Hierarchy 3	Total	Gender	Country	Age
Hierarchy 4	Total	Gender	Age	Country
Hierarchy 5	Total	Age	Country	Gender
Hierarchy 6	Total	Age	Gender	Country

Top-down and Bottom-up

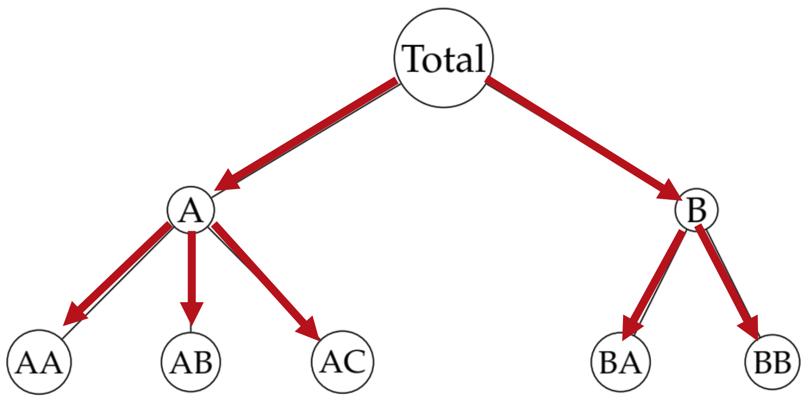
The main motivation behind the development of hierarchical forecasting has been to have consistent forecasts across levels to support decision making at different levels.

Traditionally this has been approached with the following methods:

- Top-Down (and its variants)
- Bottom-Up
- Middle-Out

Top-Down

The TD approaches requires us to forecast at the top level of the hierarchy and then disaggregate the forecasts.



Forecast here!

… and disaggregate to lower levels

There are three popular approaches to disaggregation:

- Use average historical proportions
- Use proportions of historical averages
- Forecast and use proportions of the forecasts

This is the **best**, as only this can handle seasonalities and trends appropriately.

Top-Down

With **Top-Down** we produce a forecast at the top level and then disaggregate it to the lower levels of the hierarchy.

Advantages:

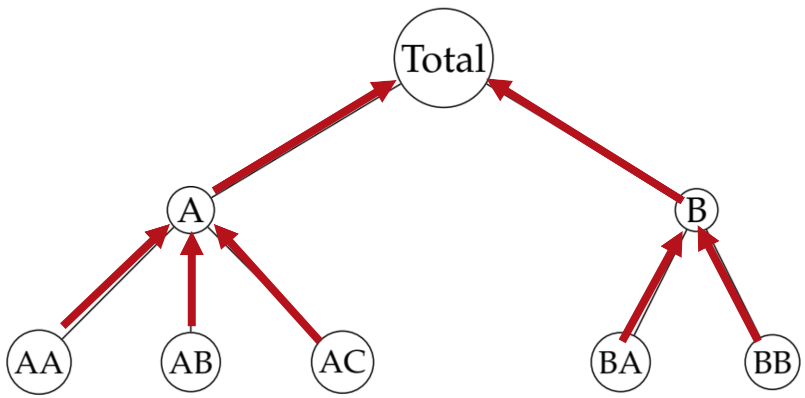
- Works well in presence of low count series (at lower/lowest levels)
- Single forecasting model easy to build
- Provides reliable forecasts for aggregate levels

Disadvantages:

- Loss of information especially at lower level time series dynamics
- Distribution of forecasts to lower levels can be difficult
- No prediction intervals

Bottom-Up

The BU approach requires us to forecast at the lowest level of the hierarchy and then aggregate the forecasts by summing them up appropriately.



Forecast here!

… and aggregate to
higher levels

Bottom-Up

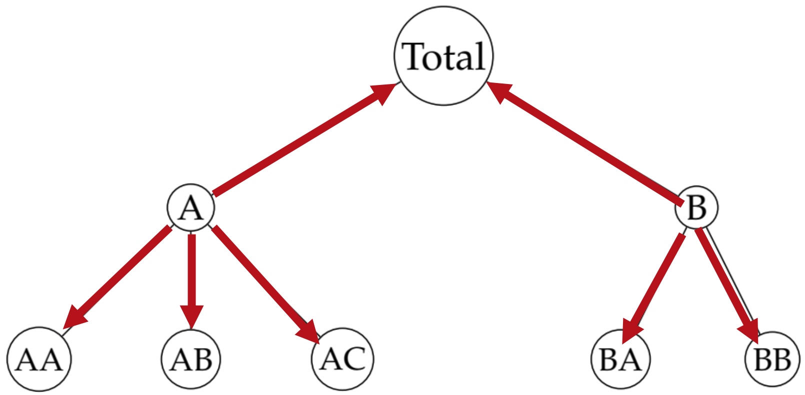
With **Bottom-Up** we produce a forecasts at the lowest level and then aggregate them to the upper levels of the hierarchy.

Advantages:
- No loss of information
- Better captures dynamics of individual (low level) time series

Disadvantages:
- Large number of time series to forecast
- Constructing forecasting model is harder because of noisy data at bottom level
- No prediction intervals

Middle-Out

The approach is a hybrid between TD and BU. We forecast at an intermediate level and (dis)aggregate as needed. The idea is to forecast at a statistical convenient level, hoping that this will be easier and more accurate.



Forecast here!

... and aggregate
and disaggregate
to other levels

If there are many different intermediate levels, there is no theoretical insight in which one to choose and this has to be demonstrated experimentally.

What to use?

There is mounting evidence against Top-Down:

- Produces biased lower level predictions, which are particularly crucial for operational decisions taken at the disaggregate levels (for e.g. inventory).
- Also, most software does not provide the best disaggregation of the forecasts, harming accuracy further.
- **But** can still be convenient when lower levels are very erratic/intermittent.

Bottom-Up often becomes the norm, as it is convenient (we do not need to determine the best Middle-Out level)

- This has the advantage that we look at the most detailed view of the data, at the cost of difficulty in modelling.
- **But** in practice most systems do some ad-hoc middle-out, as the forecast is not done at the most disaggregate level.

Important limitations

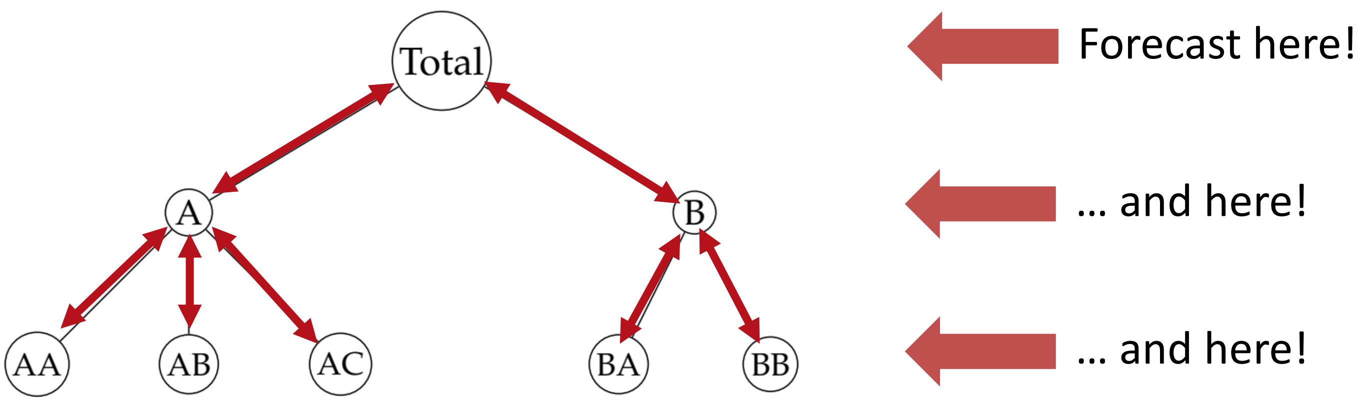
Hierarchical forecasts ensure consistent forecasts, but they come with limitations:

- We rely on a very small number of forecasts to produce predictions for the complete hierarchy – **do we trust our initial forecasts?**
- There is no guarantee that the accuracy will improve by applying any of the conventional hierarchical approaches.
- Can only handle **hierarchical** time series and cannot forecast **grouped** data. This forces us to forgo consistency across all aggregation pathways.

This has lead to the development of a new approach, the so called **optimal combinations.** This is optimal in reducing the reconciliation error while minimally changing any forecasts.

Optimal Combinations

This approach requires us to forecast at all levels and combine the predictions in a smart way.



The final prediction at each node of the hierarchy is a (linear) combination of the forecasts for the whole hierarchy, with the condition that the final forecasts are always consistent.

Optimal Combinations

Optimal combinations has several advantages over the conventional hierarchical approaches:

- It can deal both with **hierarchical** and **grouped** data.
- It has been proven theoretically that in the long term it will always be at least as good if not better than the initial forecasts → in practice that means we can expect gains in accuracy.
- It relies on combination of forecasts → has been shown to be generally beneficial, but crucially **it reduces the modelling risk**.
    - No longer rely on a few models that may be misspecified, but on as many as possible mitigating the model selection and specification risk.
- Computationally more expensive, but not prohibitive.

Optimal Combinations

Does it work in practice?

- There is consistent evidence of accuracy gains, apart from the consistency of the forecasts → both attributes improve decision making.
- For e.g.:
    - inventory decisions at store level are aligned with inventory decisions at distribution centre.
    - staffing decisions for call centres, match resources for support technicians
    - etc.
- In terms of accuracy various applications have shown gains:
    - Between 2-8% across the whole hierarchy.
    - Typically smaller gains at lowest level and larger gains at higher levels.
    - If original forecasts are very accurate, gains are small, but optimal combinations ensure consistency.

Optimal Combinations

The catch:

- This is all too new (a decade old!!!), so no commercial software offer this as a standard.
- Standard excuse: our customers do not ask for it!
  - Well, it is not the job of your customers to know innovations in forecasting!
  - But at least now you do!

Agenda

1. ~~What is hierarchical forecasting?~~
2. ~~Cross-sectional hierarchies~~
3. Temporal hierarchies

Temporal Hierarchies

Decisions need to be aligned:

- Operational short-term decisions
- Tactical medium-term decisions
- Strategic long-term decisions

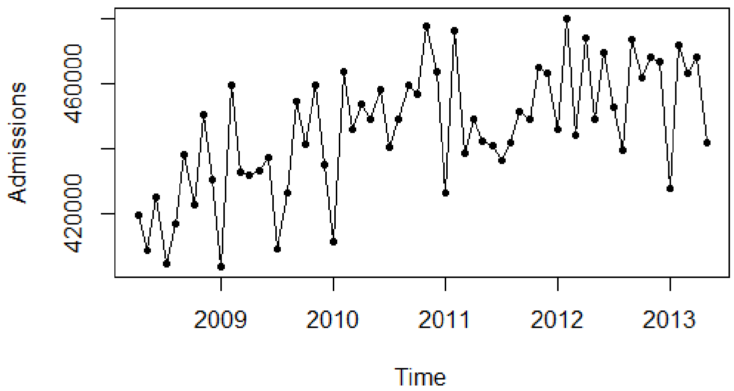
Shorter term plans are **bottom-up** and based mainly on **statistical forecasts** & expert adjustments.
Longer term plans are **top-down** and based mainly on **managerial expertise** factoring in unstructured information and organisational environment.

Given different sources of information (and views) forecasts will differ → plans and decisions not aligned.

Coherent forecasts across planning horizons can lead to less waste & costs, agility to take advantage of opportunities.

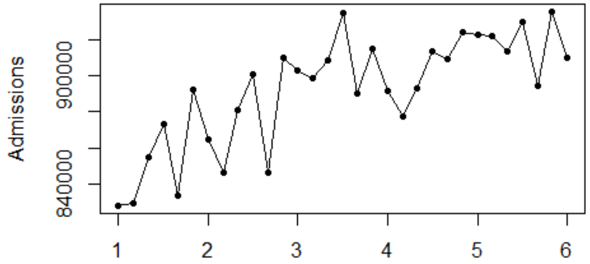
Temporal Aggregation

Consider some historical monthly sales series:

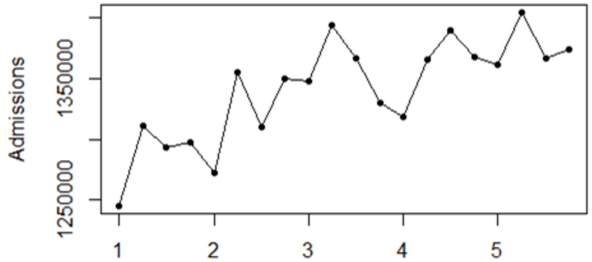


If wanted a long term forecast, we could either produce multi-step ahead forecasts, or aggregate the data and produce single-step ahead forecasts for the long horizon directly:

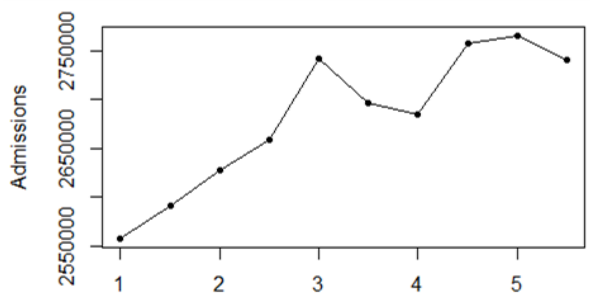
- 12 monthly forecasts vs. 1 yearly!



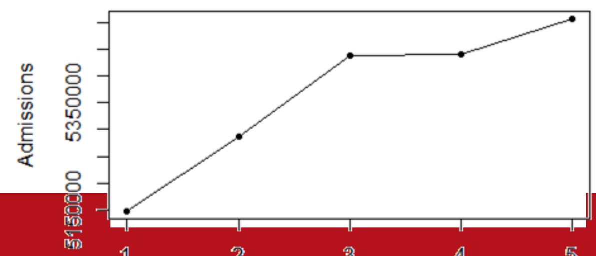
Bi-monthly



Quarterly



Half-annually



Annually

Temporal Aggregation

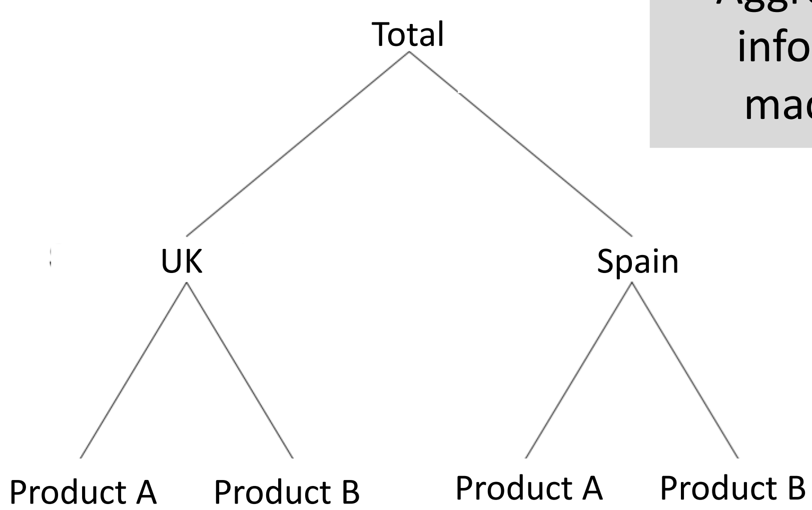
- Produce long term forecasts with multi-step predictions is risky: forecast errors accumulate!
- Temporal aggregation can help to reduce the length of the forecast.
- What does temporal aggregation do to our data?
  - at an aggregate level trend/cycle is easy to distinguish.
  - at a disaggregate level high frequency elements like seasonality and promotions typically dominate.

- Arguably both disaggregate and aggregate are useful. We can look at both and connect them in a hierarchical way.

Temporal Hierarchies

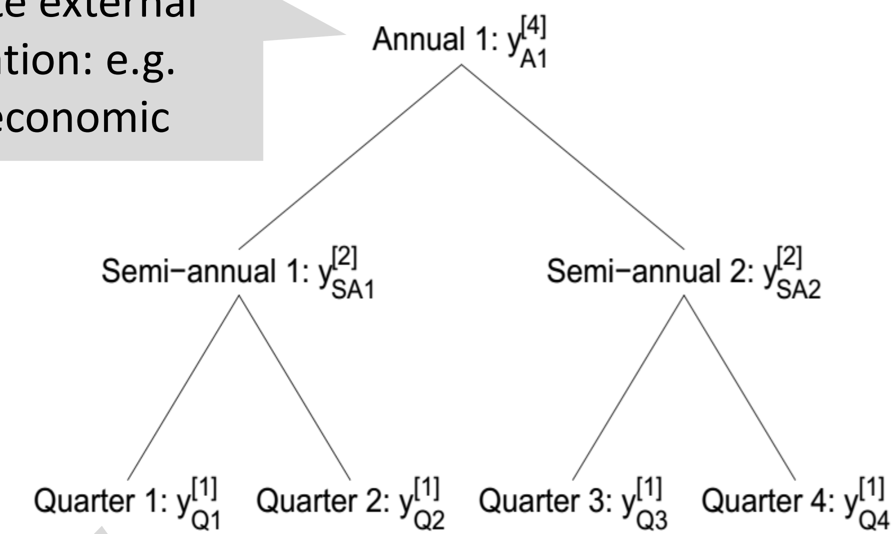
- In a hierarchical forecasting thinking we can observe that:

Cross-sectional hierarchy



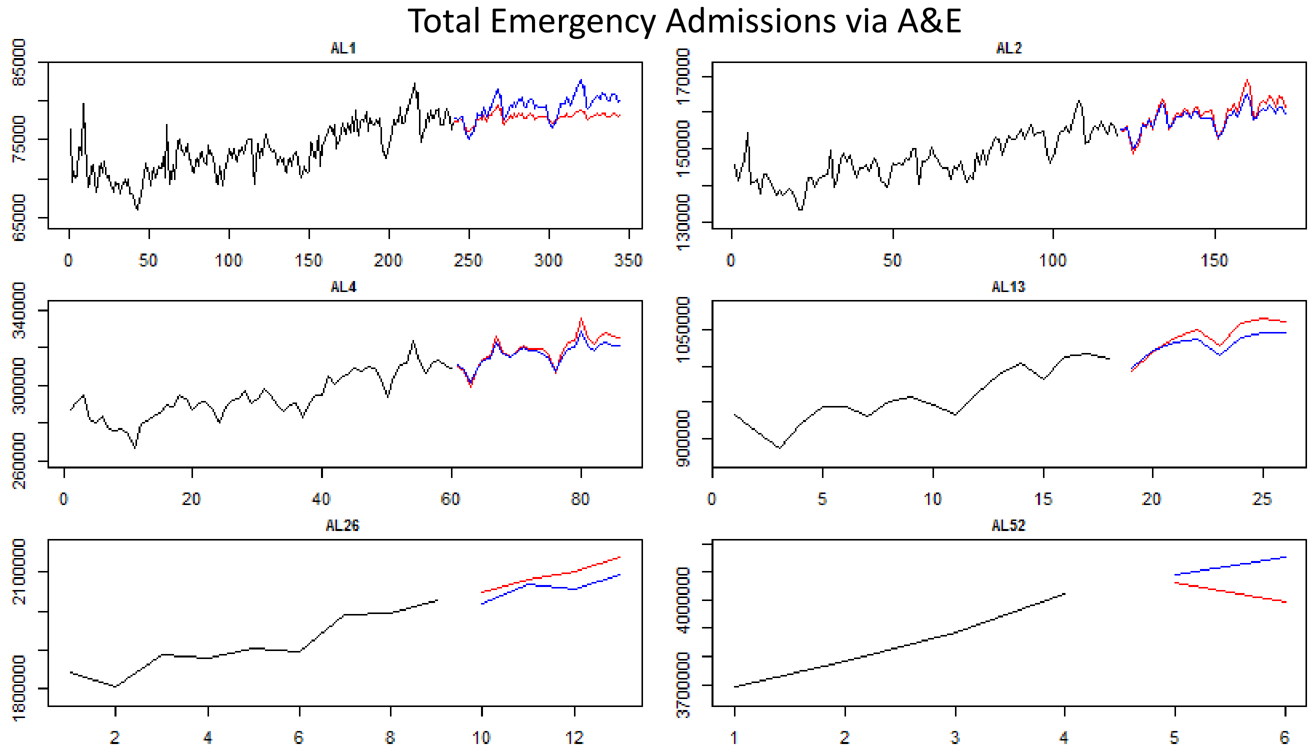
Temporal hierarchy

Aggregate external information: e.g. macroeconomic



Disaggregate internal information: e.g. promotions

Application: Predicting A&E admissions



Red is the prediction of the base model – at each level separately
Blue is the temporal hierarchy forecasts

Observe how information is `borrowed' between temporal levels. Base models for instance provide very poor weekly and annual forecasts

Application: Predicting A&E admissions

Data level	Horizon	Accuracy Change
Weekly	1	+17.2%
Weekly	4	+18.6%
Weekly	13	+16.2%
Weekly	1-52	+5.0%
Annual	1	+42.9%

- Accuracy gains at all planning horizons
- Crucially, forecasts are reconciled leading to aligned plans
- We can go one step further: merging location & temporal level predictions together

Temporal Hierarchies

What are the advantages of Temporal Hierarchies?

- They align decision making across different planning horizons.
- They are for free, i.e. they do not require any extra data from conventional forecasting.
- They have been shown to be at least as good as conventional forecasting, but typically offer accuracy gains.
- They mitigate modelling risk: the same data are modelled using alternative views. If one is poorly modelled, this is compensated by the other views. Conventional forecasting does not do that.

Cross-Temporal Hierarchies

Naturally, one can combine **cross-sectional** and **temporal** hierarchies to achieve:

- Aligned decisions across parts of the business (products, segments, markets, etc.) and horizons (operational, tactical, strategic).
- Forces information sharing.
- Further accuracy gains and mitigation of modelling risk.

Cross-sectional

- Reconcile across different items.
- Units may change at different levels of hierarchy.
- Suppose an electricity demand hierarchy: lower and higher levels have same units. All levels relevant for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU are useful. Weekly sales of organisation are not! Needed at different time scale.

Temporal

- Reconcile across time units/horizons.
- Units of items do not change.
- Consider our application. NHS admissions short and long term are useful for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU is useful for operations. Yearly sales of a single SKU may be useful, but often not!
- Operational → Tactical → Strategic forecasts.

Conclusions

- The motivation behind hierarchical forecasting has been to achieve forecast consistency to facilitate decision making.
- Conventional approaches (Top-Down, Bottom-Up, Middle-Out) have multiple limitations and more crucially there is little theory to drive their setup, **BUT** they do the job and are widely available in software.
- Optimal combinations are very useful as they can achieve all **forecast consistency**, **mitigation of modelling risk** and **gains in accuracy**.
- Temporal hierarchies is an innovative way to forecasting that enables **consistency across planning horizons** and **gains in accuracy**, particularly in the long term.
- As both cross-sectional and temporal hierarchies are cast in the same mathematical framework, it is relatively easy to combine them to cross-temporal hierarchical forecasts → one (consistent) forecast for the whole organisation.

Adoption ready?

- **Multiple Aggregation Prediction Algorithm (MAPA)**
  - Kourentzes, N.; Petropoulos, F. & Trapero, J. R. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting,* **2014***, 30*, 291-302
  - **R package on CRAN: MAPA**
  - Papers (+ additional ones), code and examples available on my website (http://nikolaos.kourentzes.com)
- **Hierarchical (cross-sectional) forecasting**
  - Hyndman, R. J.; Ahmed, R. A.; Athanasopoulos, G. & Shang, H. L. Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis,* **2011***, 55*, 2579-2589
  - **R package on CRAN: hts**
- **Temporal Hierarchies**
  - Athanasopoulos, G.; Hyndman, R. J.; Kourentzes, N. & Petropoulos, F. Forecasting with temporal hierarchies. *European Journal of Operational Research*, **2017**, *262*(1), 60-74.
  - **R package on CRAN: thief**
  - Also look at posts summarising research at: http://kourentzes.com/forecasting/2017/04/27/multiple-temporal-aggregation-the-story-so-far-part-i/

Thank you for your attention!
Questions?

Nikolaos Kourentzes (@nkourentz)

email: nikolaos@kourentzes.com

blog: http://nikolaos.kourentzes.com